

다항식 신경회로망에 의한 오존농도 예측모델

김태현^{**}, 김성신^{*}, 이종범^{**}, 김신도^{***}, 김인택^{****}, 김용국^{**}

*부산대학교 전기공학과, **강원대학교 환경학과, ***서울시립대학교 환경공학과, ****명지대학교 전기정보공학부

Modeling of Ozone Prediction System using Polynomial Neural Network

T.H.Kim*, S.S.Kim*, J.B.Lee**, Y.K.Kim**, S.D.Kim***, I.T.Kim****

*Department of electrical Engineering,Pusan National University

**Department of Environmental Science,Kangwon National University

***Department of Environmental Engineering,The University of Seoul

****School of Electrical and Information Engineering,Myong Ji University

Abstract - In this paper, we present the modeling of ozone prediction system using polynomial neural network. The Polynomial Neural Network is a useful tool for data learning, nonlinear function estimation and prediction of dynamic system. The mechanism of ozone concentration is highly complex, nonlinear, nonstationary.

The proposed method shows that the prediction to the ozone concentration based upon a polynomial neural network gives us a good performance for ozone prediction with ability of superior data approximation.

1. 서 론

최근 국내외를 막론하고 대기오염의 주된 관심사는 하계 중 대류권의 지나친 오존의 고농도 현상이다. 일반적으로 오존은 질소산화물 및 탄화수소류 등의 오염 물질의 광화학적인 작용과 일사량, 풍속, 기온 등의 기상학적인 변수들의 상호작용으로 생성되며, 지구에 존재하는 오존의 10%정도가 지표에서 10km이내인 대류권에 존재하고 있다. 오존은 주로 하계인 6월부터 8월 사이에 집중적인 고농도 현상을 보이고 있으며, 강력한 산화력으로 인해 적당량의 오존은 인간에게 이로우나 기준농도 이상의 오존은 인체의 건강이나 농작물의 수확 및 생태계에 피해를 주고 있다. 따라서 우리나라에서는 산업발달과 자동차의 증가로 인해 각종 오염물질의 배출이 증가함에 따라 지표오존의 고농도현상이 해마다 늘어나는 추세에 있고, 또한 오존은 단 시간 내에 저 농도에서 고 농도현상을 보이기 때문에 고농도 발생으로 인한 오존의 피해를 줄이기 위해 대기 오염 예보제 실시와 함께 기상학적 및 광화학적 특성에 의한 오존농도변화 및 정확한 고농도 오존 예측에 대한 연구가 진행되고 있다. 기존의 고농도 오존 예측을 위한 방법으로 통계적 방법에 의한 선형 회귀 모델(8) 그리고 다변량 통계분석(6)에 의한 방법과 신경회로망을 이용한 예측모델(2)에 의한 연구가 이루어져 왔으나, 기상학적인 변수와 오존 오염물질 간의 강한 비 선형성과 대류권내에서의 오존생성에 관한 매우 복잡한 반응기작으로 인하여 고농도 오존에 있어서는 대부분의 예측 결과가 실제 오존농도보다 낮게 예측하는 경향을 보이고 있다. 이러한 결과는 기존의 오존 농도 예측 모델링이 비선형적, 비정상적인 경향의 적절한 반영이 이루어지지 않았거나, 예측에 절대적 영향을 미치는 데이터의 부족, 모델링에 필요한 입력요소를 적절히 선택하지 못한 경우로 나눌 수 있다.

따라서 본 연구에서는 이러한 문제점을 해결하고 정확하고 신속한 고농도 오존 예측을 위하여 자료학습 및 예측 그리고 비 선형 시스템 모델링등에 사용되고 있는 다항

식 신경회로망(Polynomial Neural Network)을 이용하여 고농도 오존과 오염물질 및 기상학적 변수간의 관계를 모델링함으로써 기존의 방법에 의한 고농도 오존 예측 결과보다 우수한 예측결과를 실제 측정한 자료를 이용한 모델링을 통해서 보이고자 한다.

2. 다항식 신경회로망

2.1 다항식 신경회로망의 기본구조

다양의 관측자료와 변수들로부터 시스템의 모델을 구성하기 위해 GMDH(Group Method of Data Handling)를 이용한 다항식 신경회로망은 복잡한 동적인 시스템의 모델링과 지능제어에 응용되어지고 있다. 일반적인 형태의 다항식 신경회로망의 구조를 그림 1에 나타내었으며, 다항식 신경회로망의 입력변수와 출력과의 관계는 다음과 같은 함수식으로 표현된다.

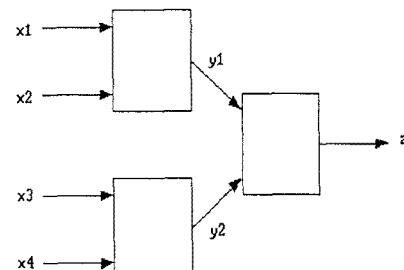


그림 1. 다항식 신경회로망의 형태

Fig 1. The basic structure of Polynomial Neural Network

여기서 각 노드에서의 출력 y_1, y_2 는

$$y_1 = w_{01} + w_{11}x_1 + w_{21}x_2 + w_{31}x_1x_2 \quad -(1)$$

$$y_2 = w_{02} + w_{12}x_3 + w_{22}x_4 + w_{32}x_3x_4 \quad -(2)$$

로 나타내어지며, 최종 출력 z 는 각 노드 출력 값의 선형결합으로 나타내어진다.

$$z = w_{03} + w_{13}y_1 + w_{23}y_2 + w_{33}y_1y_2 \quad -(3)$$

$$\begin{aligned} z = & v_0 + v_1x_1 + v_2x_2 + v_3x_3 + v_4x_4 + v_5x_1x_2 + \dots \\ & + v_{10}x_3x_4 + v_{11}x_1x_2x_3 + \dots \end{aligned}$$

$$+ v_{14}x^2x^3x^4 + v_{15}x^1x^2x^3x^4 \quad - (4)$$

각 노드의 파라미터 w_{ij} ($i=0, 1, 2, \dots, n$, $j=0, 1, 2, \dots, k$)를 결정하기 위해 Least Square Method를 사용하여 목적함수인 실제 측정된 값과 학습된 출력된 값과의 차이를 최소화하는 파라미터를 구하게 된다.

$$J = \sum_{k=1}^n (a(k) - \hat{a}(k))^2 = \| \alpha - \Phi w \|^2 \quad - (5)$$

$$w = (\Phi^T \Phi)^{-1} \Phi^T a \quad - (6)$$

이렇게 해서 구해진 파라미터가 현재노드의 출력함수를 구성하고, 다음 층 노드의 입력 값으로 들어가게 되며 이러한 과정을 반복하여 최종적으로 우수한 성능을 보이는 출력 값을 나타내는 함수를 얻을 수 있다.

3. 예측 모델

대기오염물질의 시간에 따른 변화량 예측은 관측된 변수들과 그 변수들로 구성된 예보시스템 모델의 정확성에 의하여 결정된다. 예측모델의 구성을 위하여 물리화학적인 현상의 수학적 표현을 이용할 수 있으나 대기의 현상은 매우 복잡한 동적인 시스템으로 본 연구에서는 대기현상과 오염물질에 대한 측정된 데이터로부터 동적인 다항식 신경회로망을 이용하여 예측모델을 구성하여 성능을 확장시키고자 한다.

3.1 입력변수의 선정

일반적으로 오존의 농도예측을 위해 사용되는 대기오염자료로서는 오존, 일산화탄소, 이산화질소, 아황산가스, 부유먼지등이 사용되며, 기상자료로서 풍속, 풍향, 기온, 일사량, 강수량, 습도 및 운량등이 사용된다. 이러한 자료중에서 100ppb이상의 고농도 오존에서의 강수량은 대부분 0mm이기 때문에 고농도 오존예측에 거의 영향을 미치지 못하며, 또한 풍향과 운량은 수치화의 어려움, 그리고 아황산가스와 부유먼지는 대기오염규제로 인해 그 배출량이 줄어들고 있기 때문에 입력자료에서 삭제하였다. 따라서 학습될 변수로 오존, 일산화탄소, 이산화질소, 기온, 일사량, 습도, 풍속등을 사용하였으며, 이중 이산화질소는 오전의 자료를 사용하고 나머지 자료는 하루 중 최고 오존농도가 나타나는 시간대인 12시부터 17시까지의 자료를 사용하였다. 대상자료는 서울지역중 하계에 고농도 오존이 자주 발생하는 지역인 쌍문동과 성수동 지역의 96년과 97년도 6월달부터 8월달사이의 자료를 이용하였고, 결측데이터는 모두 삭제하였다. 그림 2. 와 그림 3.에 97년도 성수동과 쌍문동 지역의 하계중 100ppb 이상의 오존농도가 발생한

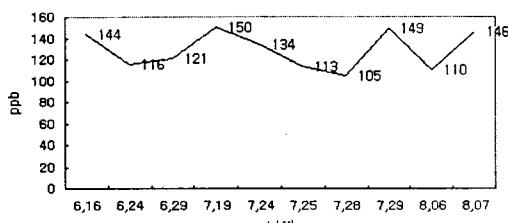


그림 2. 97년도 쌍문동 오존 농도변화(100ppb이상)

Fig 2. Variation of ozone concentration for 1997 in Ssang Mun Dong

날을 나타내었다.

3.2. 모델 구조

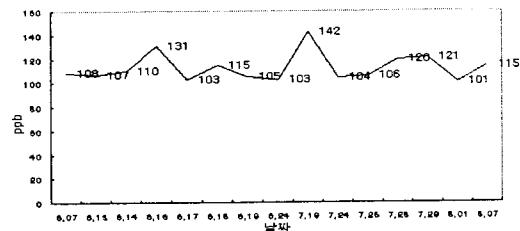


그림 3. 97년도 성수동 오존농도 변화(100ppb이상)

Fig 3. Variation of ozone concentration for 1997 in Sung Su Dong

위에서 선정한 8개의 기상자료와 오염물질자료에 대해서 오존의 농도를 출력으로 두고 나머지 7개의 변수를 입력으로 하여 학습을 시킨 결과 선택된 입력변수들과 모델의 네트워크 구조를 아래 그림 3.에 나타내었다

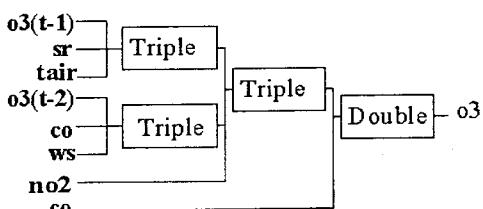


그림 4 .선택된 입력변수와 모델의 구조1

(100ppb이상, 데이터의 갯수: 81개, o3:오존, ws:풍속
tair:기온, sr:일사량, no2:이산화질소, co:일산화탄소)

Fig 4. Selected input variable and model structure(Number of data : 81)

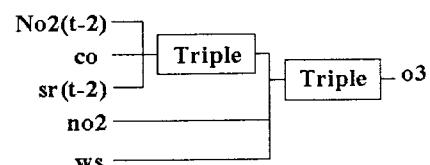


그림 5 선택된 입력변수와 모델구조2

(120ppb이상, 데이터의 갯수: 35개, ws:풍속
sr:일사량, no2:이산화질소, co:일산화탄소)

Fig 5. Selected input variable and model structure(Number of data : 35)

자료 학습시 하나의 변수에 대해 그림 4.는 81개의 자료를 사용하여 모델링 하였고, 그림 5.는 35개의 데이터를 사용하여 모델링 하였다. 두 모델 모두 100ppb와 120ppb이상의 고농도 오존의 데이터이며 데이터의 90%를 학습에 사용하였다. 그림 4.와 그림 5.에서 학습할 데이터의 갯수의 차이에 따라 입력변수와 모델의 구조가 달라짐을 알 수 있다. 고농도의 오존 예측 모델에 있어서 농도 120ppb이상의 데이터의 부족으로 학습의 적절한 효과를 기대할 수 없고, 따라서 우수한 예측 성능을 나타내지 못한다.

4. 예측 결과

다항식 신경회로망에 의해서 모델링된 오존예보 시스템은 학습 자료의 대부분이 농도 100ppb이상의 자료를 사용하였기 때문에 저농도 오존에 대한 예측값은 크게 기대할 수 없다.

아래 그림 6.에서 오존 농도100ppb이상의 데이터와 그림 .8에 80ppb이상의 데이터를 이용한 예측결과를 나타내었다. RMSE(root mean square error)가 5.25 와 13.02로서 기준의 결과치 보다 우수한 성능을 보이고 있으며 표 1.에서 보듯이 입력데이터의 평균치에서는 더욱 우수한 예측 결과를 보이고 있다.

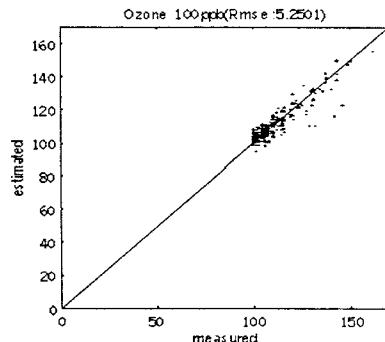


그림 6. 100ppb이상에서의 결과

Fig 6. Ozone plot over 100 ppb

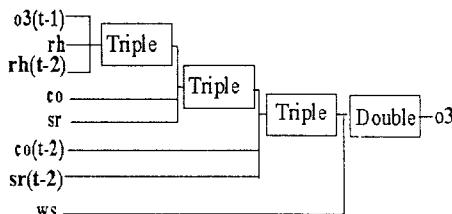


그림 7. 선택된 입력변수와 모델의 구조3

(100ppb이상, 데이터의 갯수: 173개, o3:오존
rh:상대습도,sr:일사량,ws:풍속,co:일산화탄소)

Fig 5. Selected input variable and model structure (Number of data : 173)

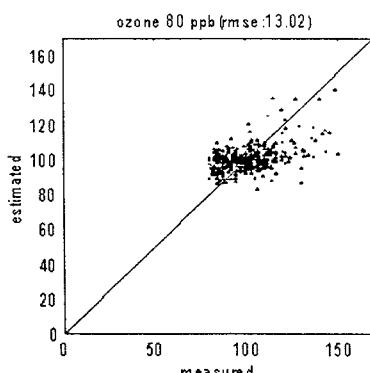


그림 8. 80ppb이상에서의 결과

Fig 8. Fig 6. Ozone plot over 100 ppb

오존농도	100-110	111-120	121-130	131-140	141-162
에러	3.6175	4.5499	7.2211	7.7065	12.7962
데이터수	103개	38개	14개	10개	8개

표 1. 농도대별 에러 분포 (100ppb이상)

Table 1. Error distribution in each ozone

concentration period(over 100ppb)

그림 6.과 8.에서 보듯이 비교적 error값은 작지만 오존 주의보 이상의 고농도 부분(120ppb이상)에서는 예측값이 실제값보다 낮게 평가되고 있다. 이는 이 농도 대에서 학습된 데이터의 갯수가 너무 적어서 적절한 학습이 이루어지지 않았음을 의미한다. 따라서 적절한 학습을 위해 또 다른 변수의 추가가 이루어져야 할 것으로 본다.

4. 결 론

정확한 오존농도의 예측은 오존의 피해가 가시화되는 현재의 시점에서 매우 중요한 영향을 미칠 수 있다.

특히 오존 경보 발생의 가능성이 농후한 지역에서는 정확한 예보가 절대적으로 중요하므로 보다 나은 예측의 필요성이 강조된다. 본 논문에서 고농도 오존예측을 위해 다항식 신경회로망(Polynomial Neural Network)을 이용하였고 실측 데이터를 이용한 모델링 결과가 우수함을 보였다. 오존 농도 예측에 있어서 우수한 예측을 위해서는 모델의 구조를 고정시키는 것 보다는 매일의 새로운 자료를 모델의 새로운 입력변수로 사용하여 예측 모델의 구조를 조정해 나가는 것이 더 우수한 예측 결과를 낼 것으로 본다. 그리고 예측결과에서 보듯이 입력된 자료의 평균 값 부분에서 모델의 성능이 대체로 우수하기 때문에 두 세가지의 복합적인 모델(저농도 혹은 고농도)을 통한 예측을 행한다면 우수한 결과가 나올 것으로 본다.

(참 고 문 헌)

- [1] A.G.Ivahnenko, "Polynomial theory of complex system, *IEEE trans. Syst. Man and Cybern.*, vol, SMC-12, pp364-378, 1971.
- [2] 김용국,이종범, "하계의 일 최고 오존농도 예측을 위한 신경망 모델의 개발", 한국 대기보전학회지, vol10. no4, pp224-232, 1994.
- [3] Duc Trung Pham and Liu Xing, "Neural Networks for Identification, Prediction and control" Springer-Verlag, 1995.
- [4] S. Chen and S.A. Billings,"Neural networks for nonlinear dynamic system modeling and identification " in Advances intelligent Control(C.J.Jarris, ed.),ch. 4, Taylor and Francis, 1994.
- [5] K.S.Narendra and K.Parthasarathy, "Identification and control of dynamical systems using neural network" *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp4-27, 1991.
- [6] 허정숙,김동술, "다면량 통계분석을 이용한 서울시 고농도 오존의 예측에 관한 연구" 한국 대기보전학회지.vol. 9,no. 2, 1993.
- [7] G.E.P.Box and G.M. Jenkins, "Time Series Analysis, Forecasting and Control, San Francisco, Holden Day, 1976.
- [8] 이선기,이종범, "서울지역의 광화학 오염농도추정을 위한 중화귀 모델의 개발" 기상연구논문집, vol. 8,no. 1, pp71-85 ,1991.